

REDES NEURAIS COM TENSORFLOW



# AGRUPAMENTOS

DIEGO RODRIGUES DSC

INFNET

# CRONOGRAMA

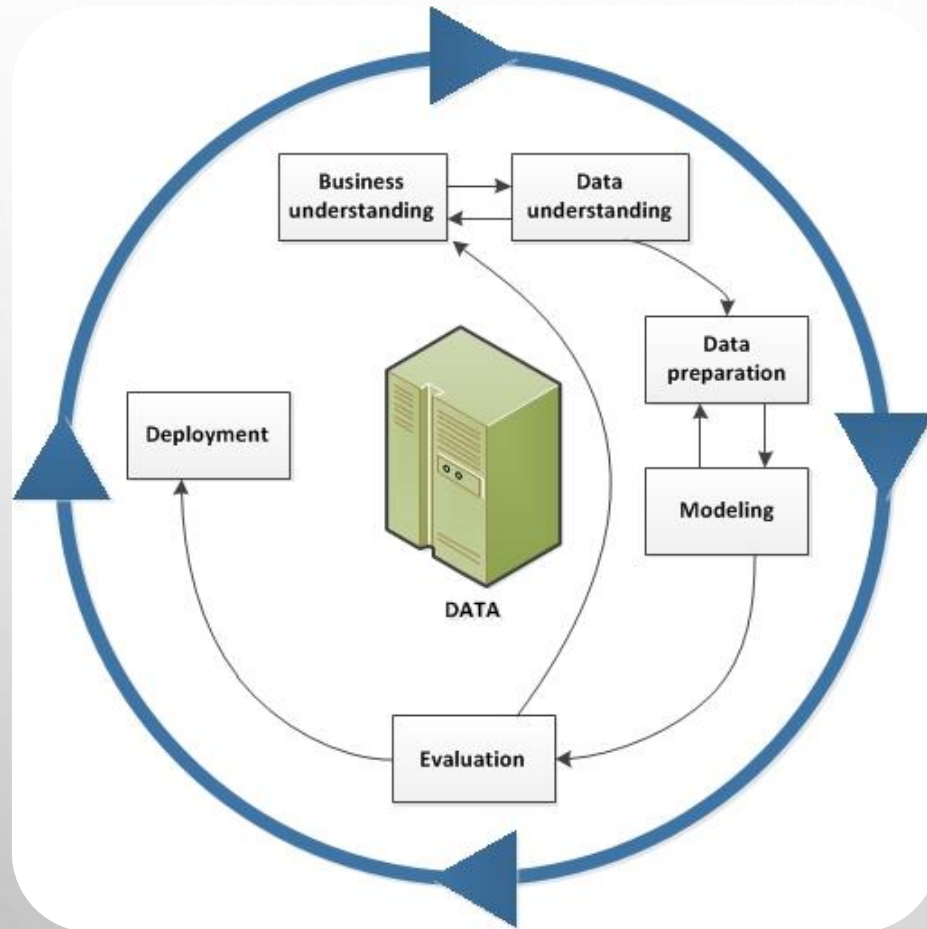
<b>Dia</b>	<b>Aula</b>	<b>Trab</b>
29/07	Perceptron de Rosenblatt	
31/07	Classificação: Neurônio Sigmóide	
05/08	Classificação: Rede Neural Feedforward	Grupos
07/08	Classificação: Rede Neural Profunda	
12/08	Regressão	base de Dados
14/08	Agrupamento	
19/08	Séries Temporais	Modelos
21/08	Apresentação dos Trabalhos Parte I	



# AGRUPAMENTOS

- PARTE 1 : MODELOS DE AGRUPAMENTO
  - BUSINESS UNDERSTANDING
  - MODELING
- PARTE 2 : PRÁTICA
  - NOTEBOOK: SOM IRIS
- PARTE 3 : TRABALHOS
  - ESCOPO & EVOLUÇÃO

# PARTE 1 : TEORIA



# CROSS INDUSTRY PROCESS FOR DATA MINING (CRISP-DM)

The image features a light gray background with a subtle gradient. In the top-left and bottom-right corners, there are clusters of realistic water droplets of various sizes, some overlapping. A faint, circular watermark is visible in the upper center of the page. The text "BUSINESS UNDERSTANDING" is centered in the middle of the page.

# **BUSINESS UNDERSTANDING**

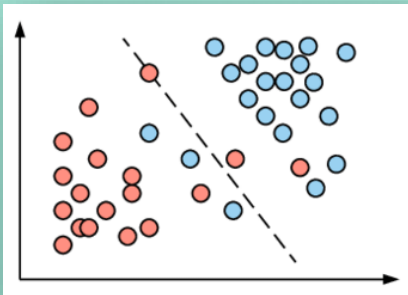
# NOVO CICLO CRISP

Algoritmo	Representação	Preparação	Modelagem	Validação
<ul style="list-style-type: none"><li>• Reta 2 Pontos</li><li>• NN 10% VAL</li><li>• NN 10 Folds</li><li>• PS10</li><li>• PS10</li><li>• SOM</li></ul>	<ul style="list-style-type: none"><li>• 2D</li><li>• 2D</li><li>• 2D</li><li>• 4D / 3 Classes</li><li>• 7D / 1D Regressão</li><li>• 4D</li></ul>	<ul style="list-style-type: none"><li>• Nenhuma</li><li>• Nenhuma</li><li>• Scale</li><li>• Scale</li><li>• Scale</li><li>• PCA</li></ul>	<ul style="list-style-type: none"><li>• Reta 2 Pontos</li><li>• 1 Neurônio</li><li>• 1 Hidden</li><li>• 1 Hidden</li><li>• 1 Hidden</li><li>• SOM 10x10</li></ul>	<ul style="list-style-type: none"><li>• Nenhuma</li><li>• Precisão/Recall</li><li>• Precisão/Recall</li><li>• Acurácia</li><li>• MSE</li><li>• MSE</li></ul>

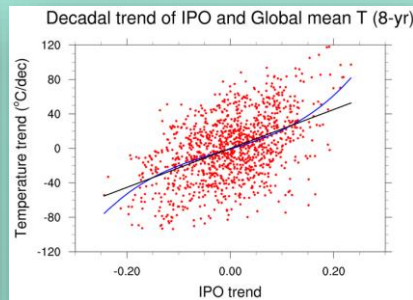
- Mapa Auto Organizável para Encontrar Grupos
- Análise da saída do mapa por atributo
- Análise da matriz U



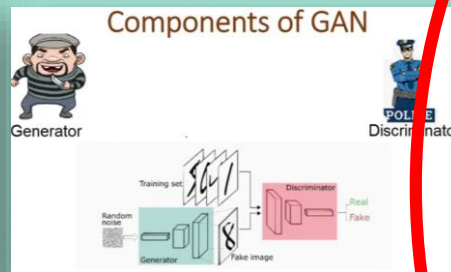
## APRENDIZADO SUPERVISIONADO



CLASSIFICAÇÃO

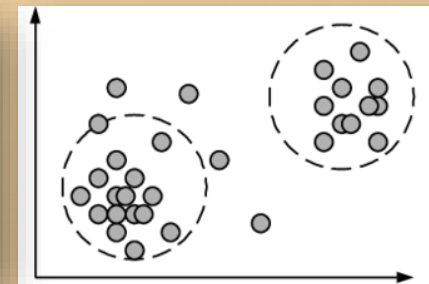


REGRESSÃO



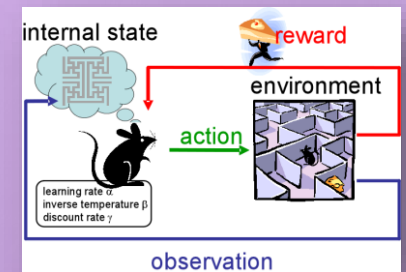
GENERATIVO

## APRENDIZADO NÃO- SUPERVISIONADO



AGRUPAMENTO

## APRENDIZADO POR REFORÇO




REFORÇO





## ○ Aprendizado Não-Supervisionado

Não existe um **conhecimento “a priori” dos grupos** contidos nos dados. Algoritmos de agrupamento dependem fortemente de uma definição de “**distância**” ou “**similaridade**” entre as observações.



# Agrupamento (Clustering)

Um bebê consegue **agrupar objetos por cor, tamanho, formato** e muitos outros atributos que ele pode observar nos objetos.

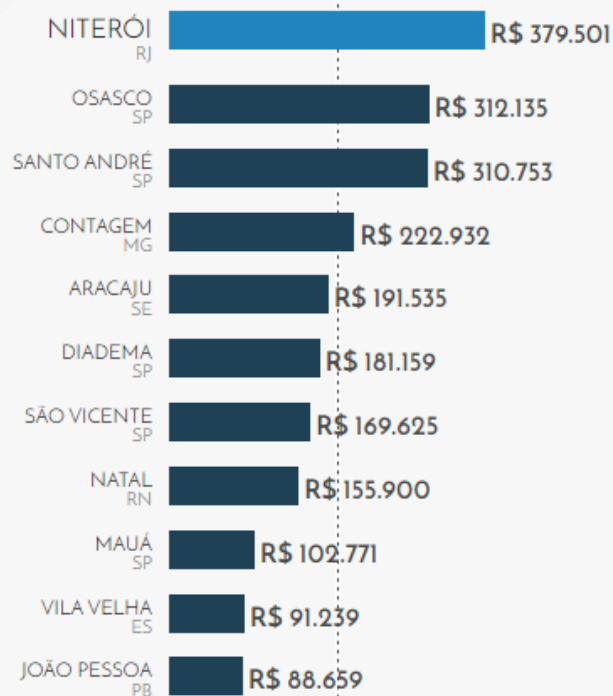
Diferentes maneiras de organizar os objetos são diferentes **estruturas de agrupamentos** existentes em uma amostra de dados.

Um **modelo de agrupamento** é usado para **identificar grupos**, ou estruturas de agrupamentos, nos dados.



De quantas maneiras estes blocos podem ser organizados em grupos?

# REPRESENTAÇÃO: COMO ENCONTRAR OS 10 MUNICÍPIOS MAIS SIMILARES A NITERÓI?



MÉDIA R\$ 200.565

## VARIÁVEIS QUE FORMAM O GRUPO - COMPARAÇÃO

Seu município	Média do grupo
Domicílios urbanos - (QTD)	
169.162	169.822
Características do Entorno	
79,34%	73,59%
Domicílios subnormais - (QTD)	
24.286	21.725
Renda média domiciliar	
R\$ 4.687	R\$ 2.503
Saneamento básico - (QTD)	
133.750	136.548

\*No gráfico ao lado, é possível comparar o município selecionado com os 10 outros municípios brasileiros de perfil mais semelhante para cada item de receita.

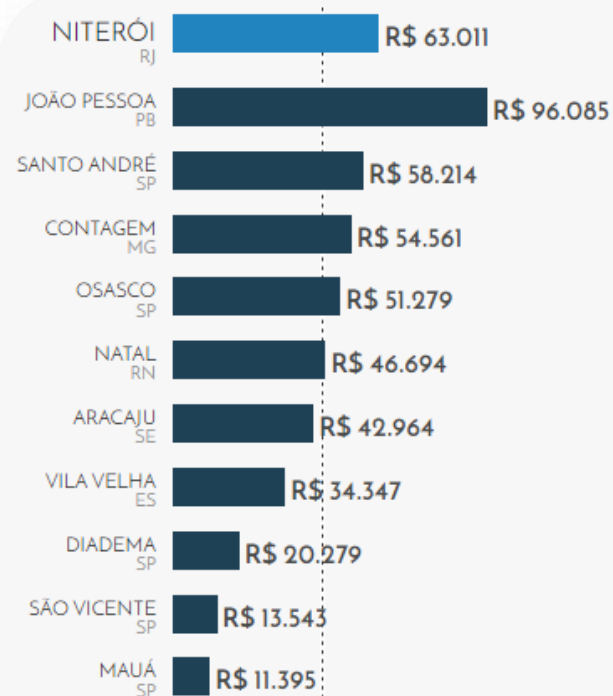
Para cada um destes, foi definido o conjunto de variáveis que mais afetam seu resultado - por exemplo, frota de veículos influencia fortemente o valor total de IPVA.

Por meio dos valores dessas variáveis, chega-se aos 10 municípios mais comparáveis com o selecionado.

Veja acima as variáveis que foram utilizadas para o componente de receita definido.

Clique em cada variável acima para entender sua importância.

entende-se que a importância de cada variável é determinada pelo seu impacto no resultado final. Por exemplo, a frota de veículos é uma variável que influencia fortemente o valor total de IPVA.



MÉDIA R\$ 44.761

## VARIÁVEIS QUE FORMAM O GRUPO - COMPARAÇÃO

Seu município	Média do grupo
Domicílios urbanos - (QTD)	
169.162	169.822
Características do Entorno	
79,34%	73,59%
Domicílios subnormais - (QTD)	
24.286	21.725
Renda média domiciliar	
R\$ 4.687	R\$ 2.503
Saneamento básico - (QTD)	
133.750	136.548

\*No gráfico ao lado, é possível comparar o município selecionado com os 10 outros municípios brasileiros de perfil mais semelhante para cada item de receita.

Para cada um destes, foi definido o conjunto de variáveis que mais afetam seu resultado - por exemplo, frota de veículos influencia fortemente o valor total de IPVA.

Por meio dos valores dessas variáveis, chega-se aos 10 municípios mais comparáveis com o selecionado.

Veja acima as variáveis que foram utilizadas para o componente de receita definido.

Clique em cada variável acima para entender sua importância.

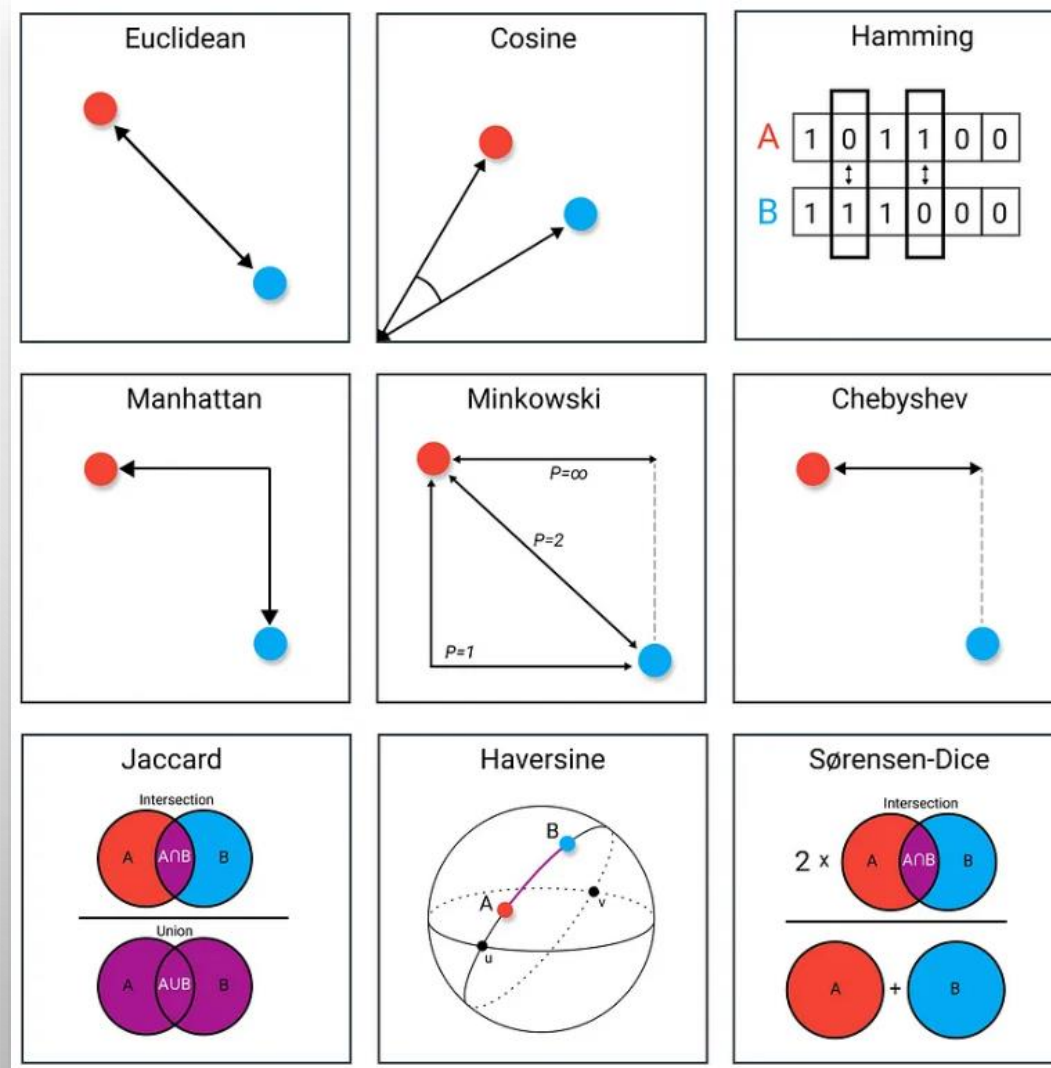
entende-se que a importância de cada variável é determinada pelo seu impacto no resultado final. Por exemplo, a frota de veículos é uma variável que influencia fortemente o valor total de IPVA.

# MODELING

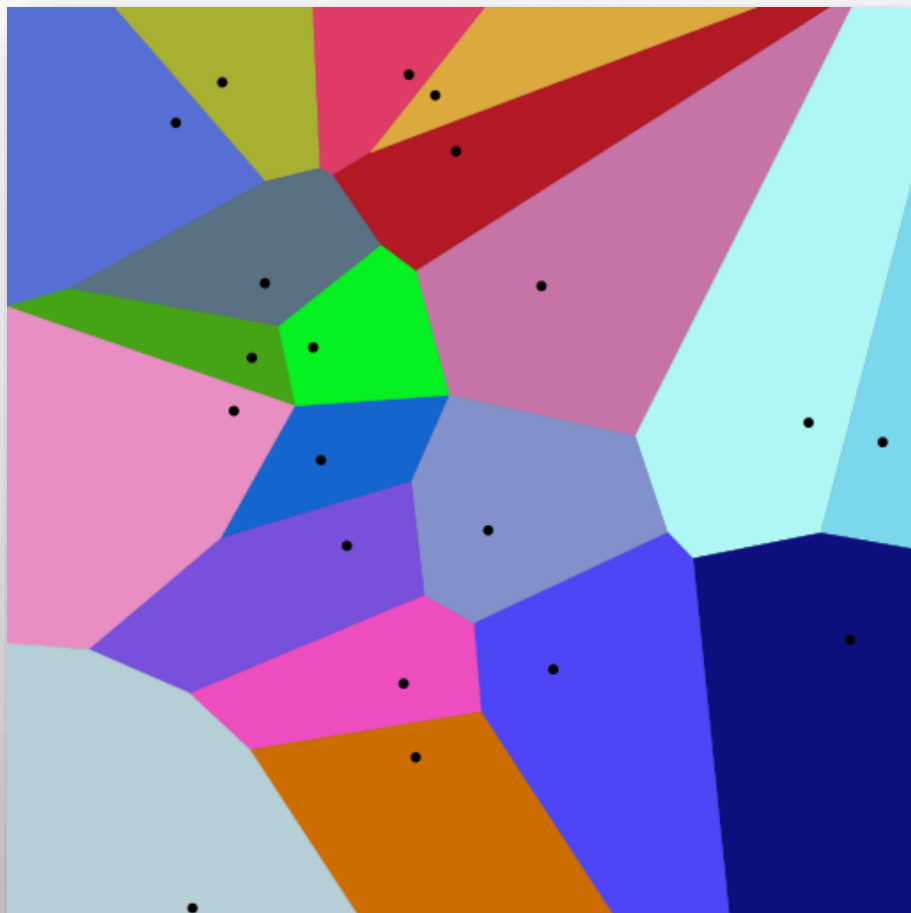


**DISTÂNCIA**

# ALGUMAS MÉTRICAS DE DISTÂNCIA







**Distância Euclideana**



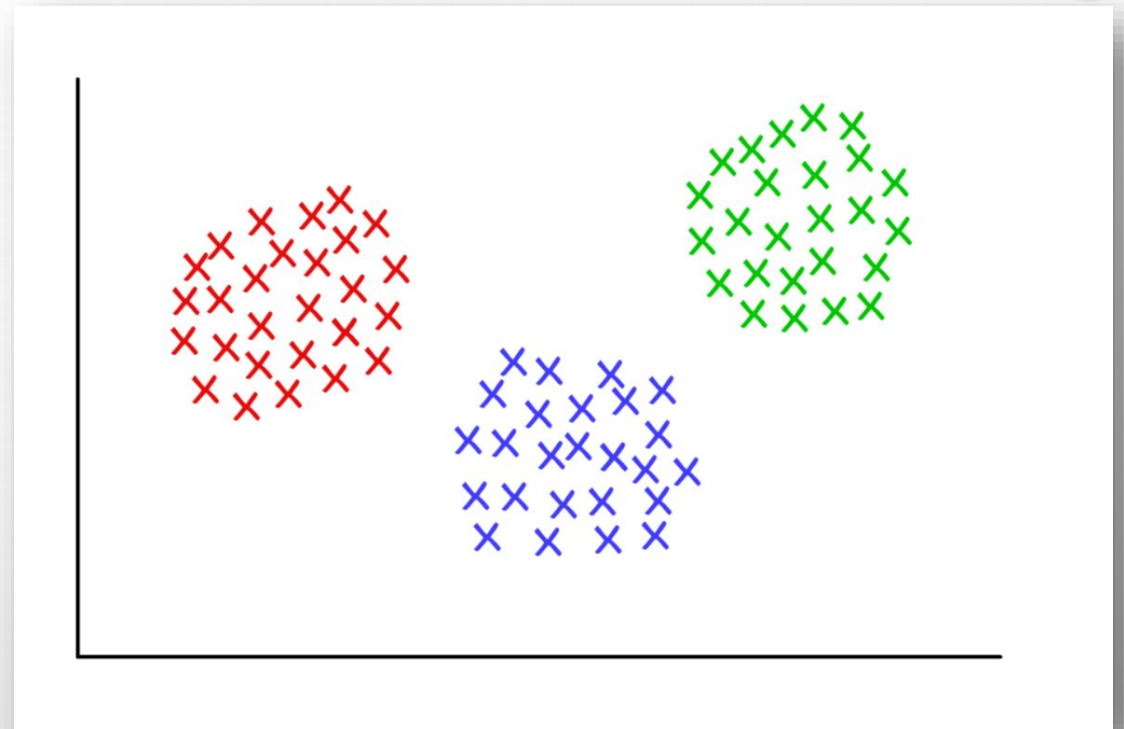
**Distância de Manhattan**

The image features a light gray background with a subtle gradient. In the top-left and bottom-right corners, there are clusters of realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. In the center of the image, the word "ALGORITMOS" is written in a bold, black, sans-serif font.

# ALGORITMOS

## AGRUPAMENTO: ALGORITMOS

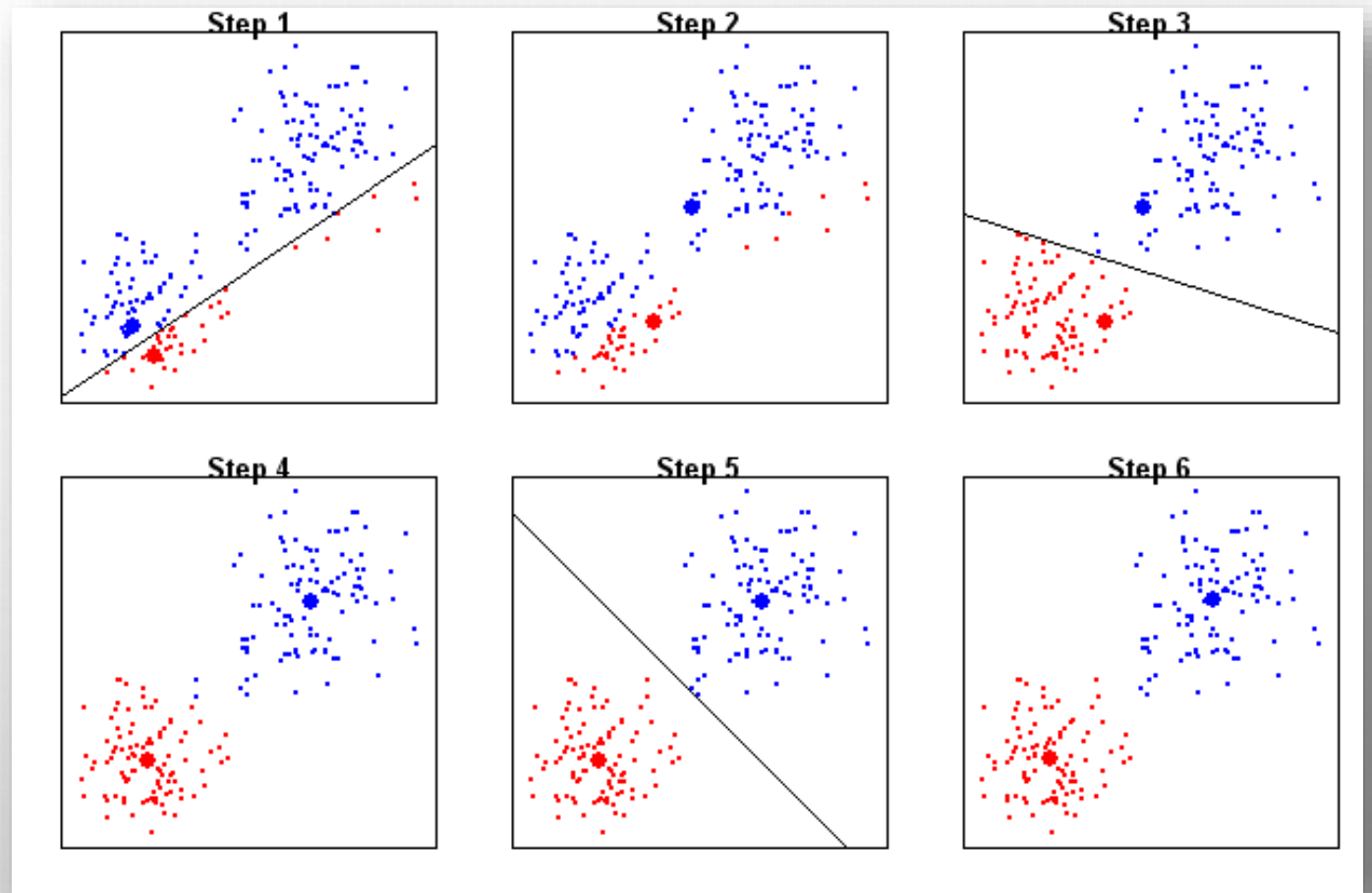
- 1) K-Means
- 2) Hierárquico
- 3) DBSCAN
- 4) Mapa Auto-Organizável



Além da escolha do algoritmo, os resultados do agrupamento dependem diretamente dos atributos e da **métrica escolhida para definir similaridade** entre os objetos.

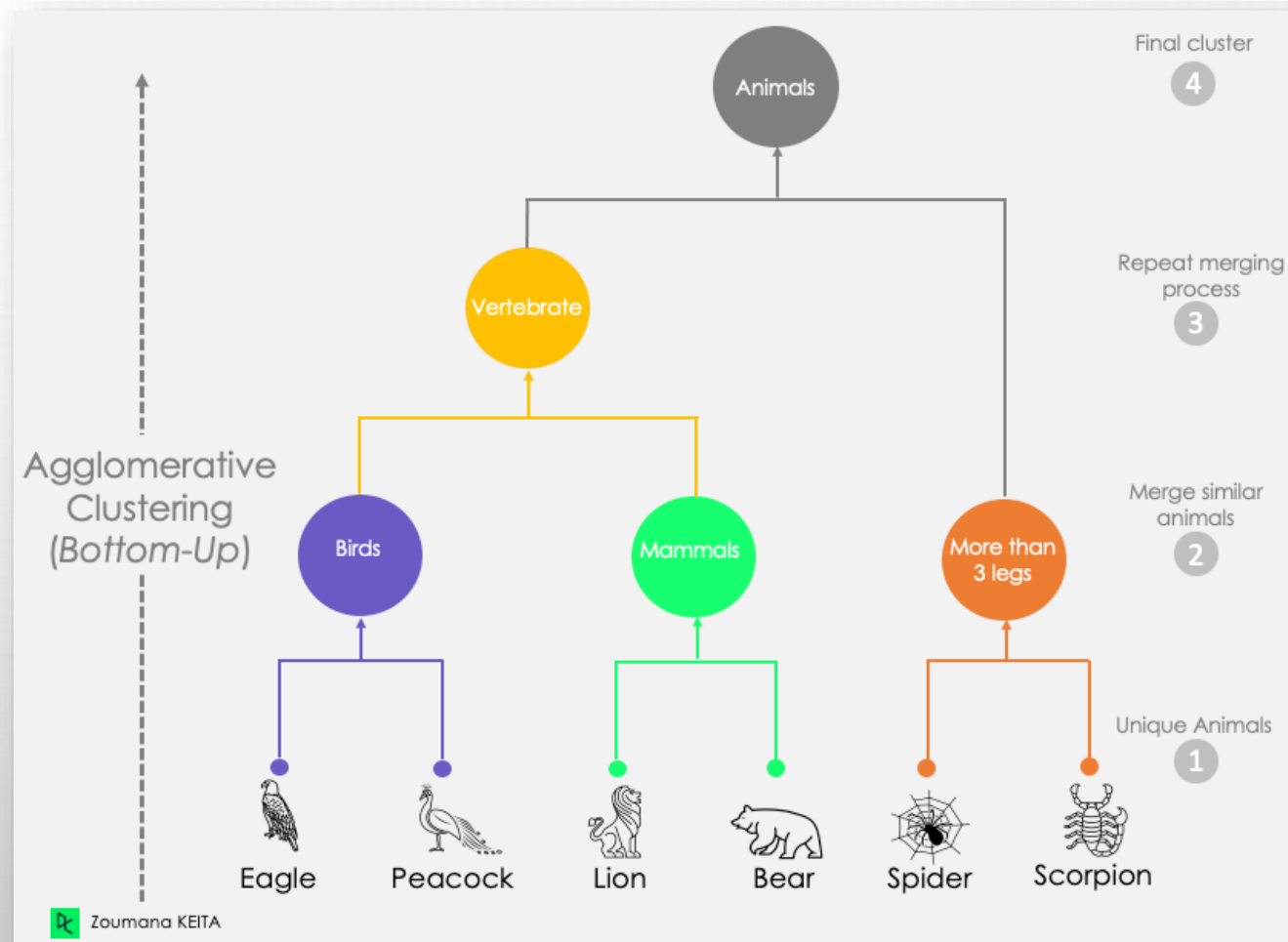
# PARTICIONAMENTO : K-MEANS

**K-means** (ou K-médias) parte de **K centroides** (centros de agrupamento) e através de iteração, **recalcula os centroides** até que particione os dados em **K grupos**.



# AGRUPAMENTO HIERÁRQUICO

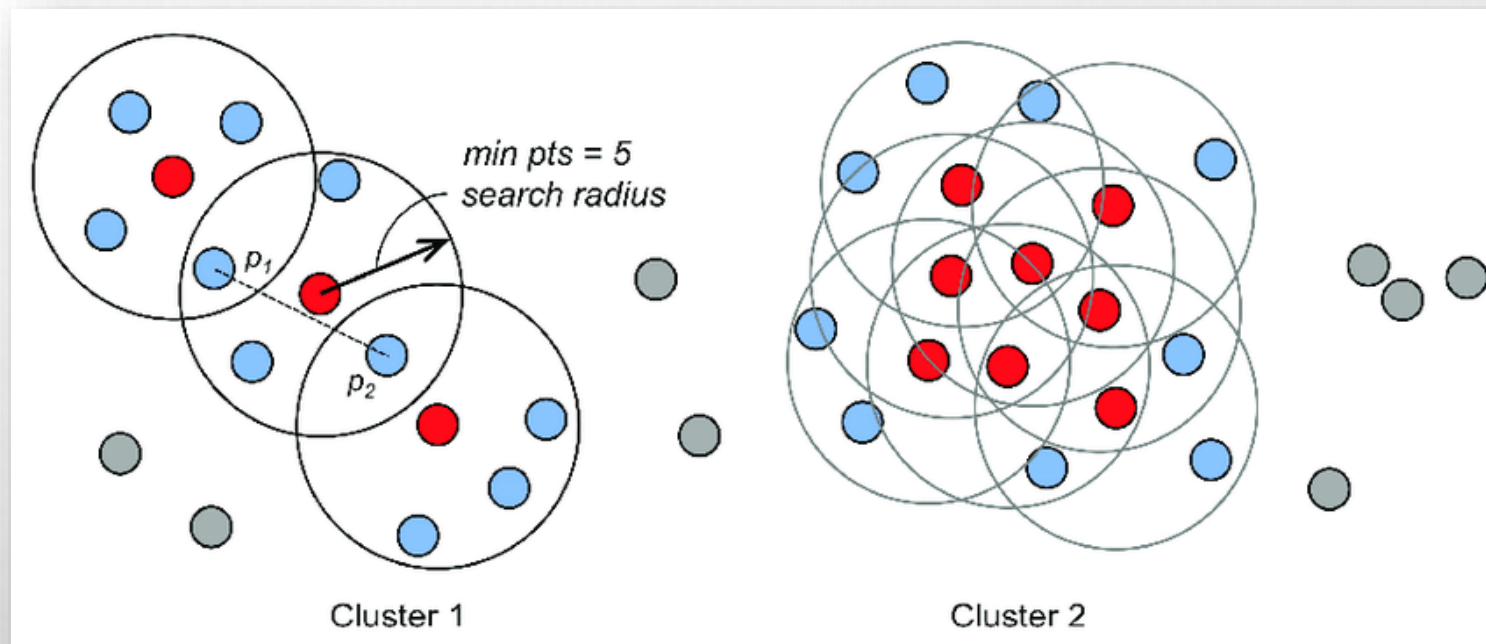
Esse algoritmo depende de mais uma definição de “**Linkage**” para decidir como **calcular similaridade entre um grupo e um indivíduo**, e permite que **diferentes arranjos de grupos** sejam detectados, variando o **limiar de distância**.



# DBSCAN : DENSIDADE

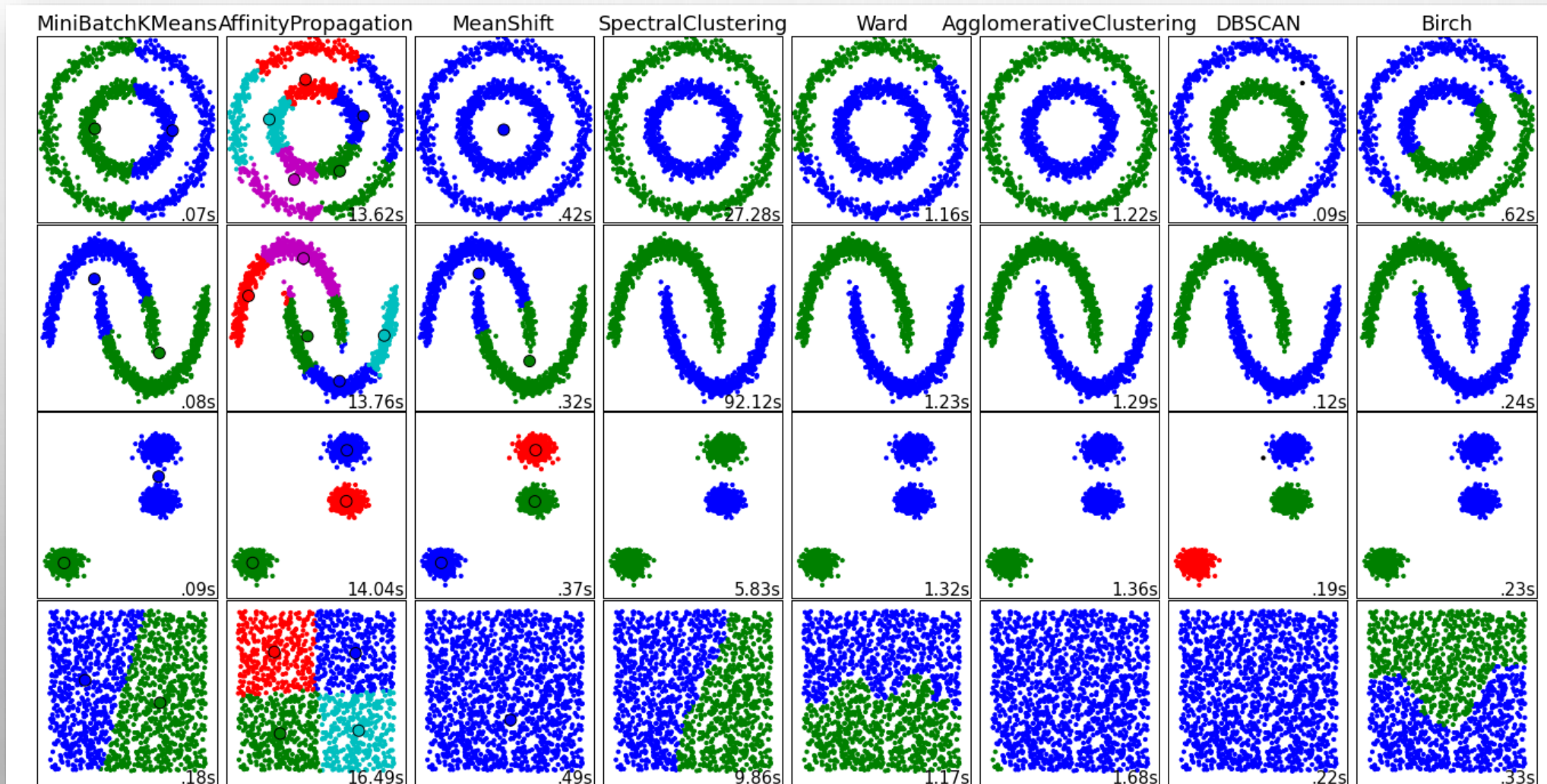
Density-Based Spatial Clustering of Applications with Noise

Baseado em uma **similaridade mínima e quantidade de vizinhos** para ser considerado um **ponto central**, DBSCAN **agrupa pontos que tenham vizinhos comuns**.

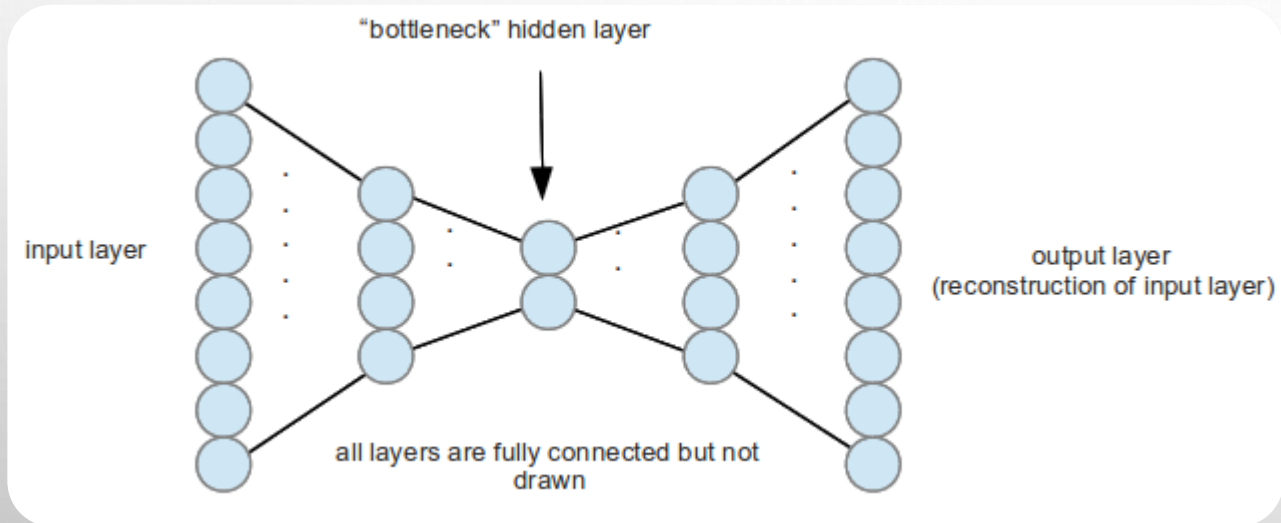




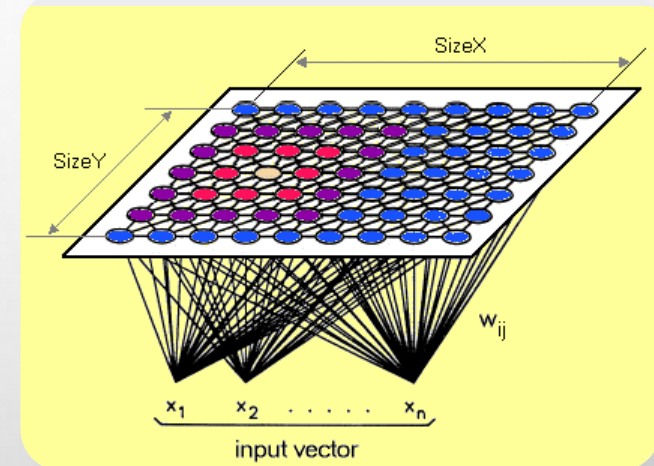
# COMPARAÇÃO DE ALGORITMOS : SKLEARN



# ○ “CLUSTERIZADOR UNIVERSAL”



Rede AutoEncoder

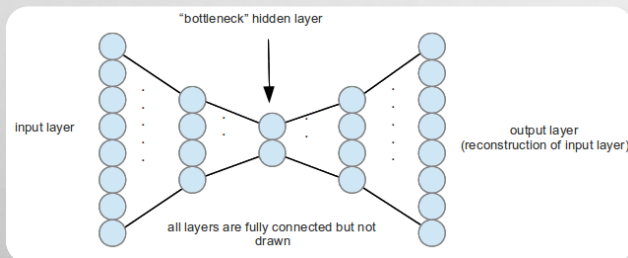


Mapa Auto  
Organizável

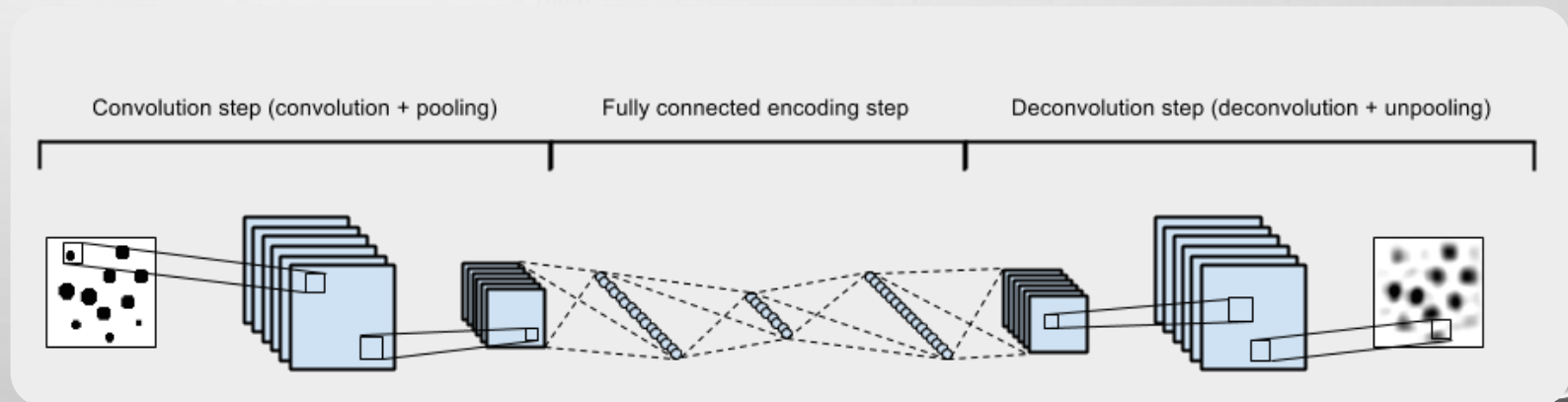
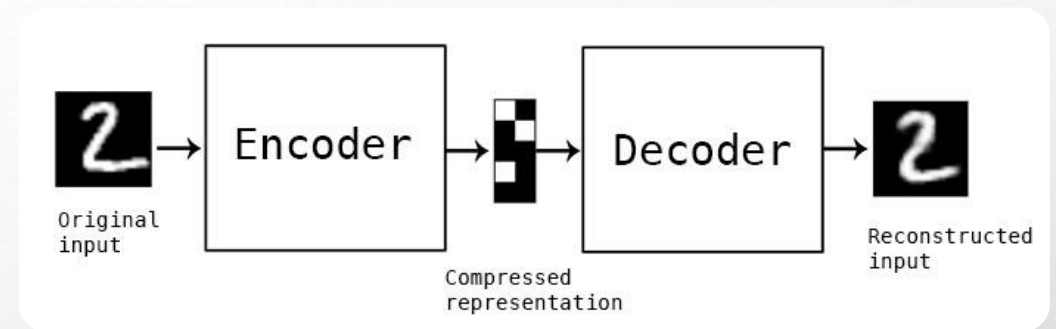
# REDE AUTO ENCODER

Encontra uma  
representação de  
menor  
dimensionalidade  
do dado

Encontra um  
hiperespaço  
reduzido contendo  
toda a informação



Rede AutoEncoder

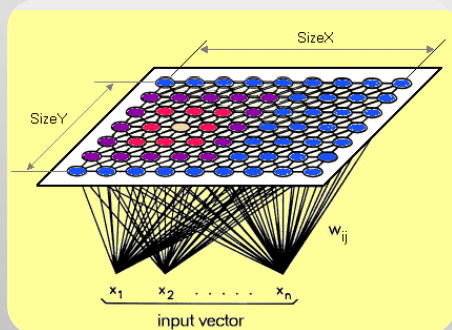




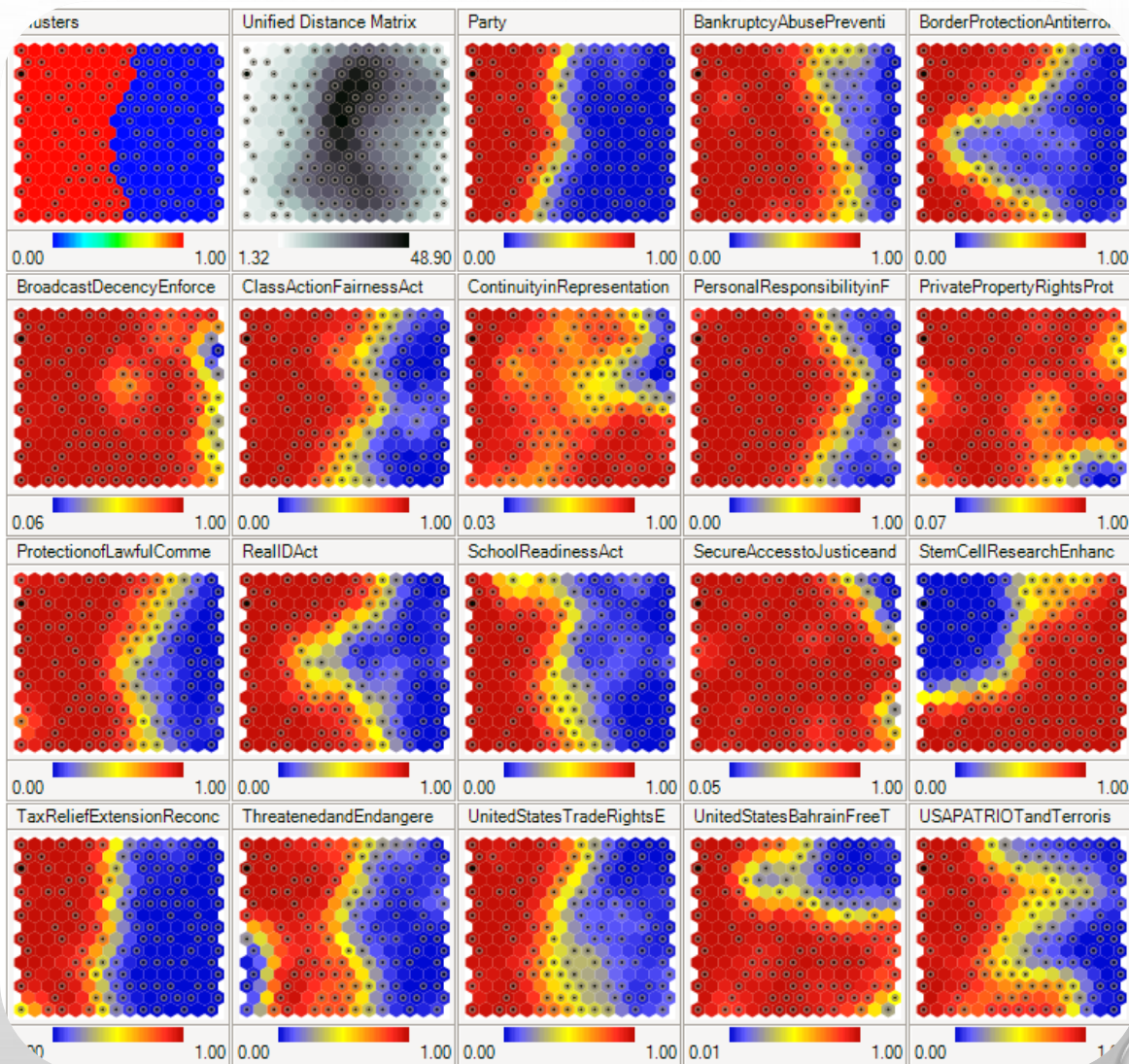
# MAPA AUTO ORGANIZÁVEL

Transforma uma  
entrada  
multidimensional  
em um mapa  
bidimensional

Cada neurônio  
serve como  
“centróide” de  
uma pequena  
região do espaço



Mapa de Kohonen



The background is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes, some overlapping. A faint, circular watermark is visible in the upper center of the page.

## PARTE 2 : PRÁTICA

# AMBIENTE PYTHON



4. Variáveis  
Aleatórias



1. Editor de Código

5. Visualização



2. Gestor de Ambiente



6. Machine  
Learning



3. Ambiente  
Python do Projeto



3. Notebook  
Dinâmico



# PROBLEMA DE NEGÓCIO

## Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



**Iris Setosa**



**Iris Versicolor**



**Iris Virginica**

Iris Setosa

Iris Versicolor

Iris Virginica

# REPRESENTAÇÃO



Iris Setosa



Iris Versicolor

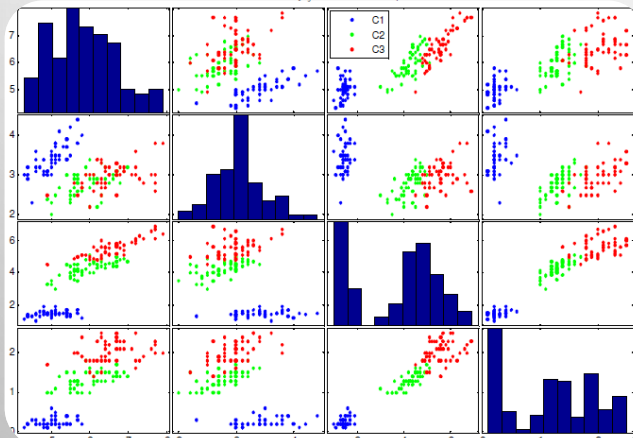


Iris Virginica

## Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



<http://archive.ics.uci.edu/ml/datasets/Iris>

Espaço de  
atributos com  
**4 dimensões!**

# MODELAGEM

## MAPA AUTO ORGANIZÁVEL

- REPRESENTAÇÃO: 4 ATRIBUTOS  $>$  2 DIMENSÕES NO MAPA
- HIPERPARÂMETROS: NÚMERO DE NEURÔNIOS DO MAPA 10X10.
- TREINAMENTO: BASE DE TREINO COMPLETA.
  - MSE



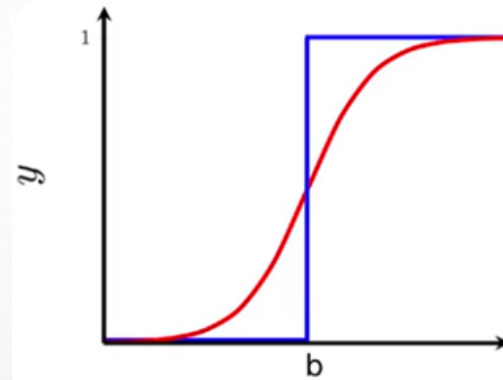
Iris Setosa



Iris Versicolor



Iris Virginica



$$y = \frac{1}{1 + e^{-(w^T x + b)}}$$

# REGRESSÃO IRIS



EXERCÍCIO: SOM IRIS

PRÓXIMA AULA: SÉRIE  
TEMPORAL

